

.*Tsp*            *Xba*I            Start codon  
 GTT TAA ATAC GCG CCG AGG TTT AAT ATG TCT GTT GCC TTG TTA TGG GTT CCT TGT CAG  
 Met Ser Val Ala Leu Leu Trp Val Val Ser Pro Cys Asp

Transit peptide of phytoene synthase  
 GTC TCA AAT GGG ACA AGT TTC ATG GAA TCA GTC CCG GAG CGA AAC CGT  
 Val Ser Asn Gly Thr Ser Phe Met Glu Ser Val Arg Glu Gly Asn Arg

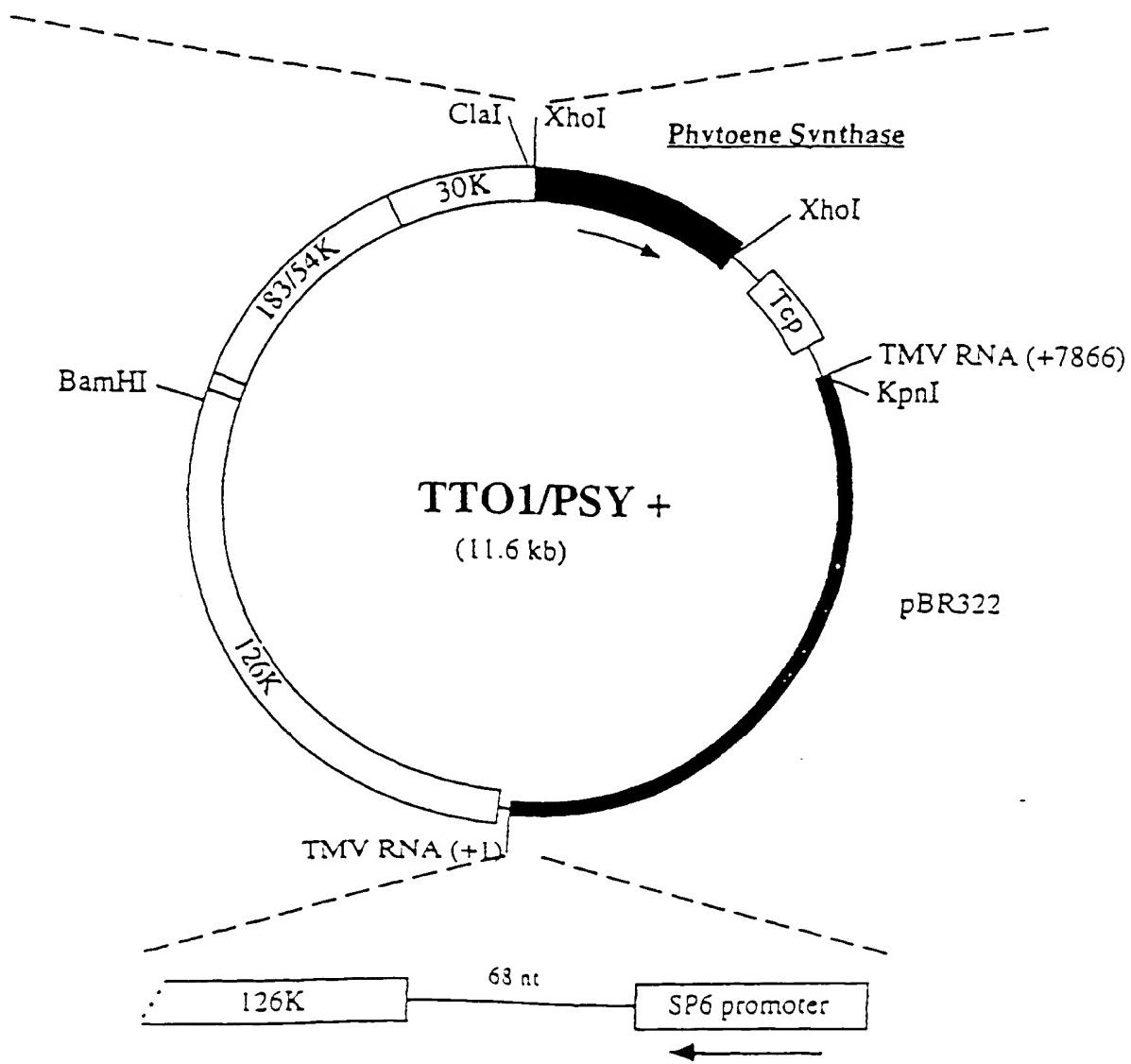
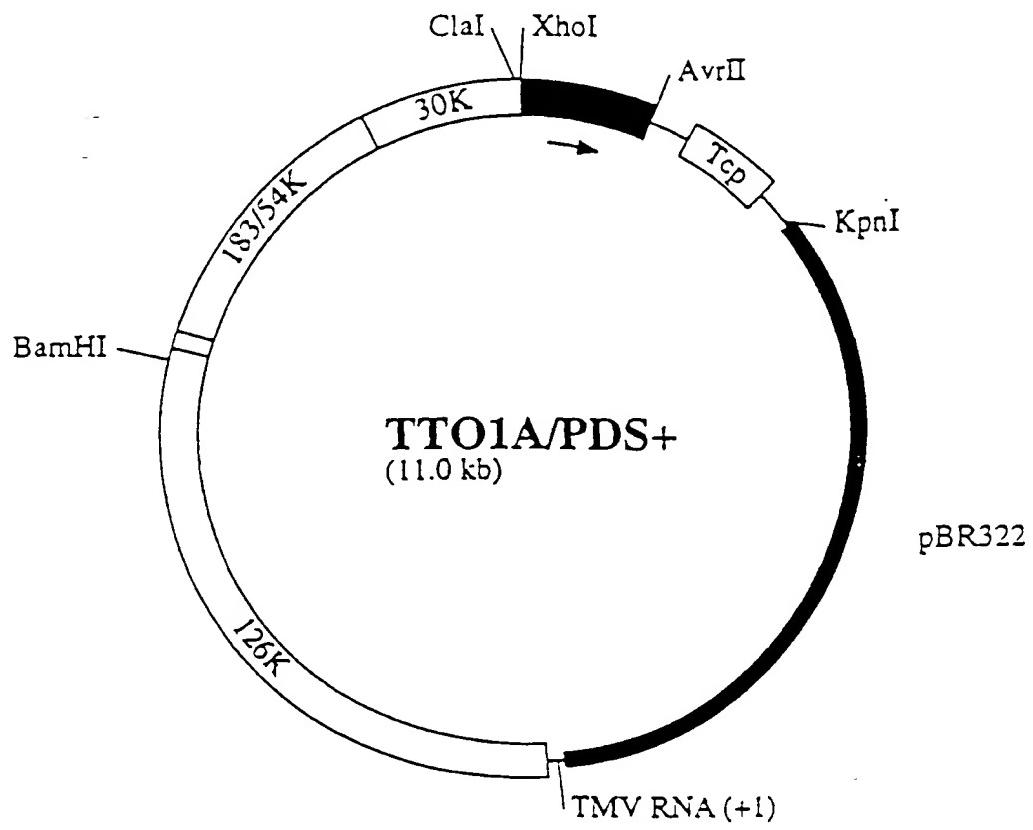


Figure 1



**Figure 2**

.tsp            XbaI            Start codon  
 GTTTTAAATACGCTCGACTGCAGC ATG GAA ACC CTT CTA AAG CCT TTT CCA TCT CCT TTA CTT  
 Met Glu Thr Leu Leu Lys Pro Phe Pro Ser Pro Leu Leu

→ TCC ATT CCT ACT CCT AAC ATG TAT AGT TTC AAA CAC AAC TTC ACT TTT  
 Ser Ile Pro Thr Pro Asn Met Tyr Ser Phe Lys His Asn Ser Thr Phe

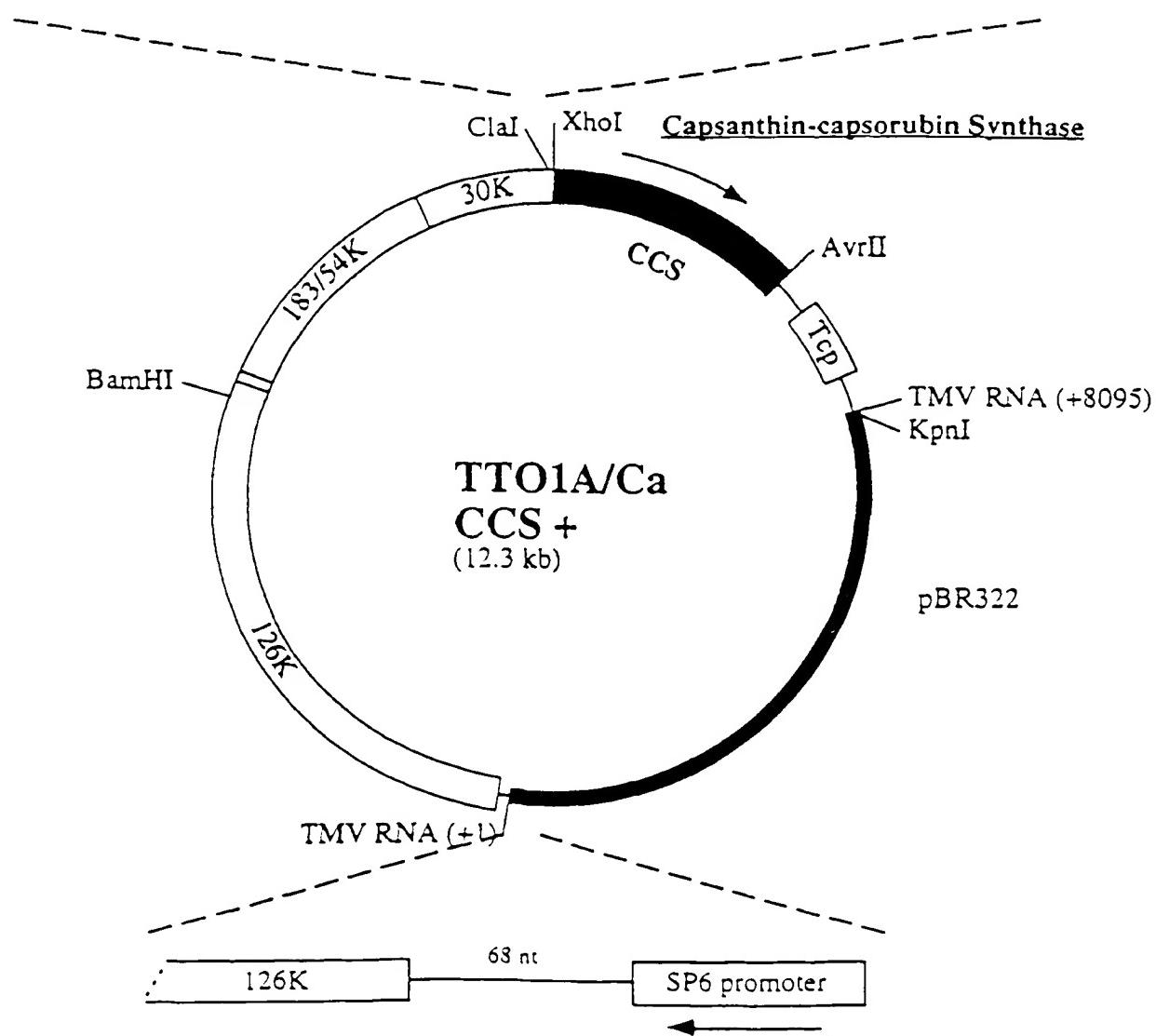
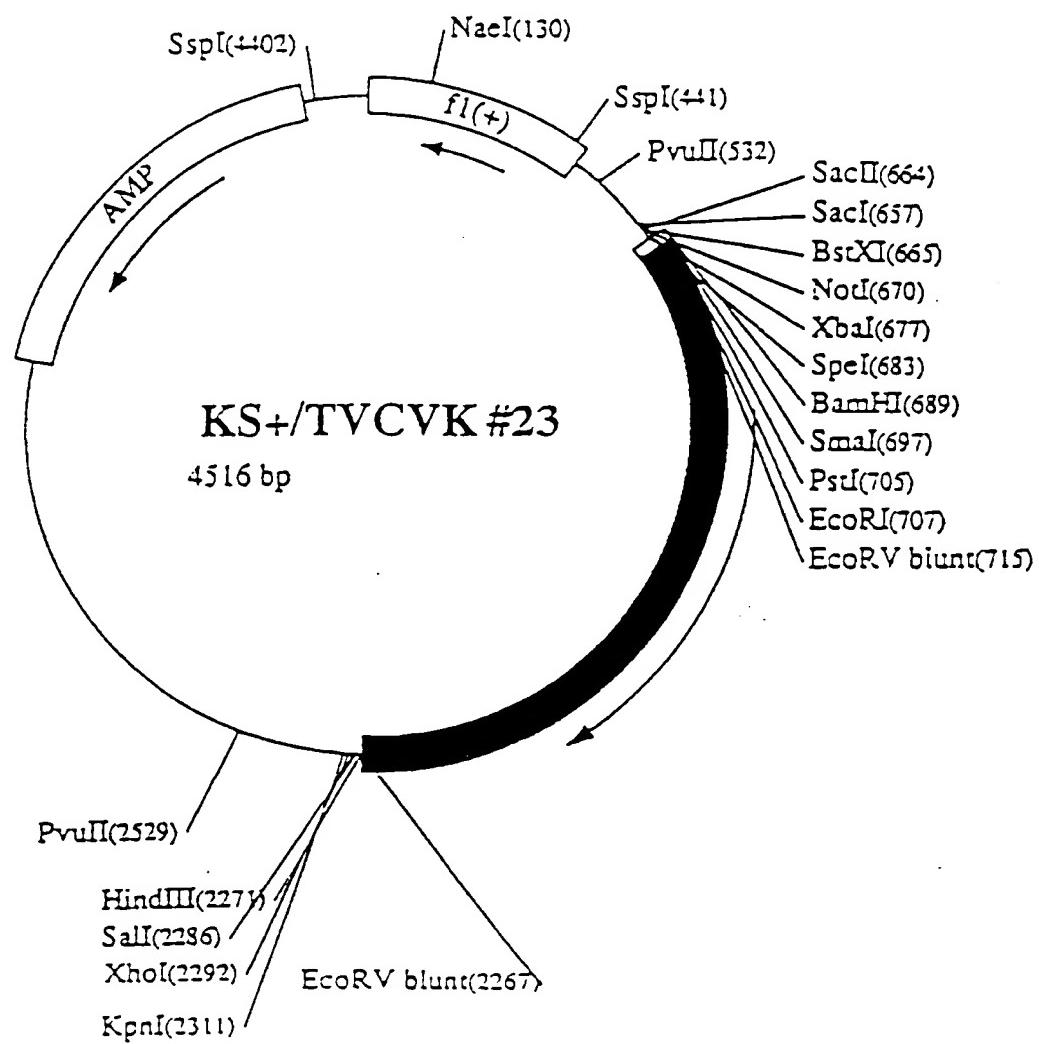


Figure 3



**Figure 4**

5'p                    XbaI      Start codon  
 GTTTAAATACGCTCGAGCC ATG GCT TCC TCA GTT CTT TCC TCT GCA GCA GTT CCC ACC CGC  
 Met Ala Ser Ser Val Leu Ser Ser Ala Ala Val Ala Thr Arg

**RUBISCO SSU Chloroplast Transit Peptide (*N. tabacum*)**

AGC AAT GTT GCT CAA GCT AAC ATG GTT GCA CCT TTC ACT GGC CTT  
 Ser Asn Val Ala Gln Ala Asn Met Val Ala Pro Phe Thr Gly Leu

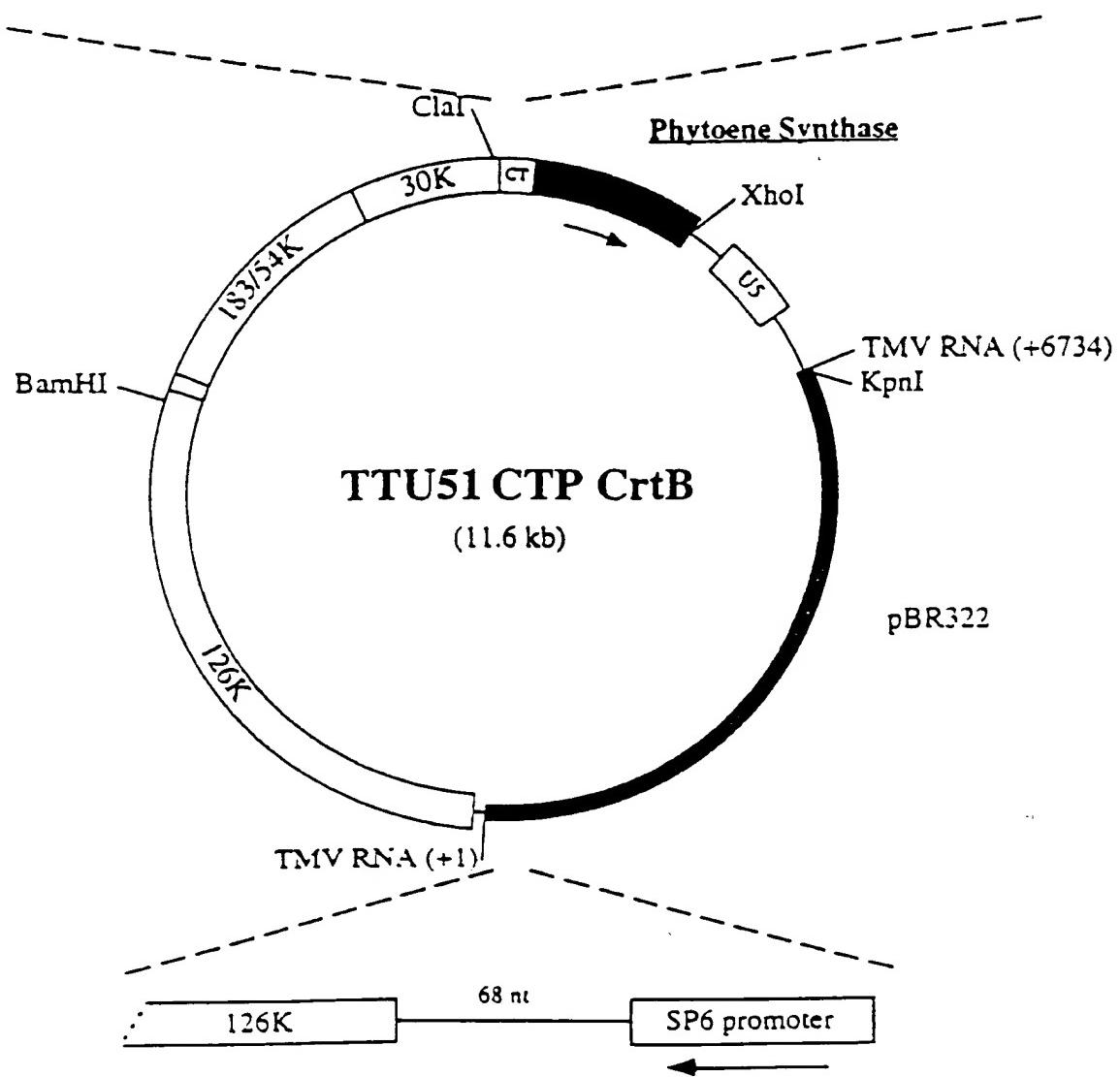
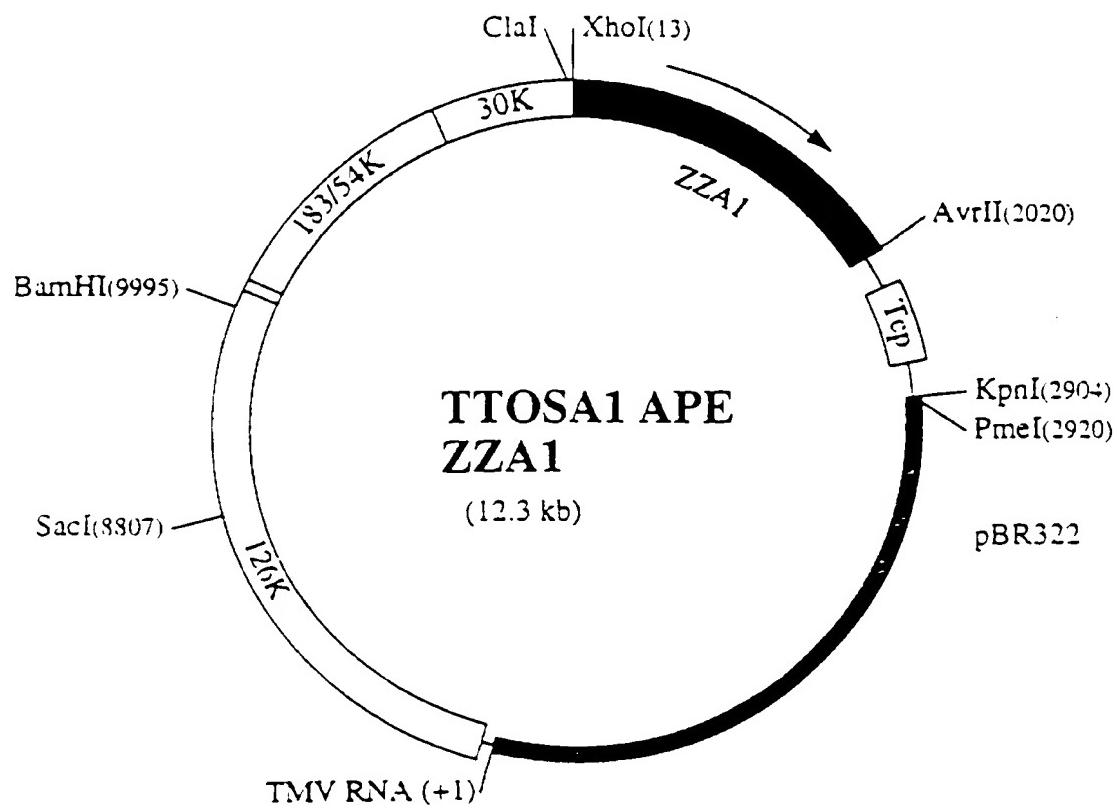
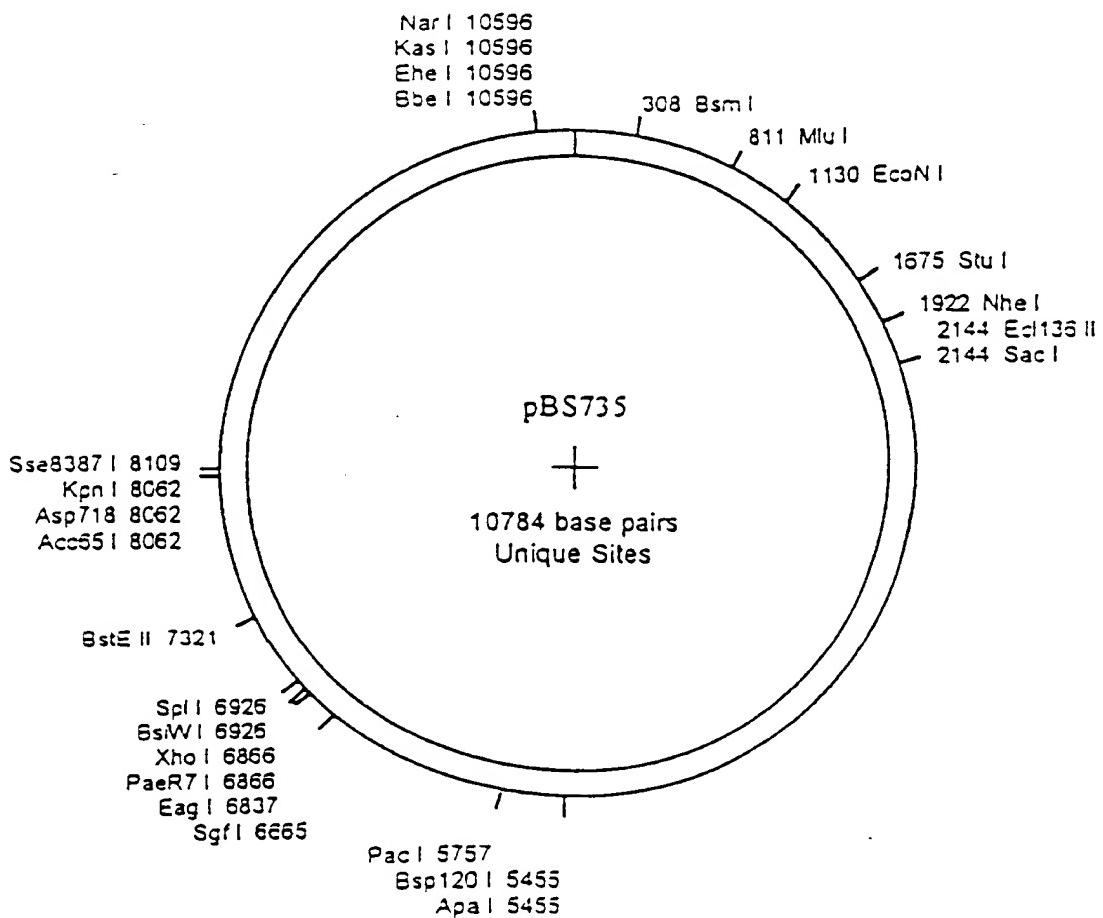


Figure 5



**Figure 6**



**Figure 7**

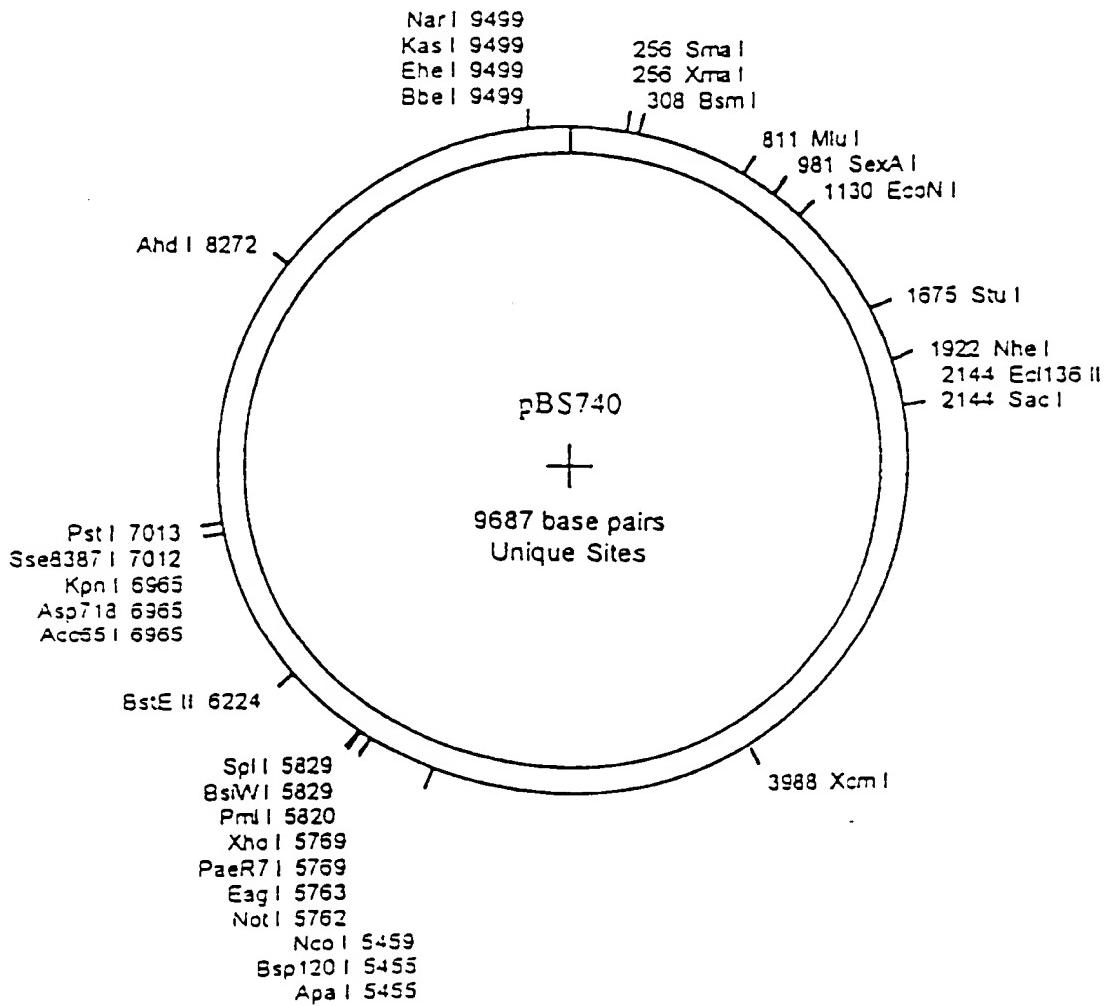


Figure 8

.usp                    XbaI     Start codon  
 STTTTAAATACCGCTGAGG ATG ATC AGA TTC TTA GTC CTC TCT TTG CTA ATT CTC ACC CTC T...  
 Met Ile Arg Phe Leu Val Leu Ser Leu Leu Ile Leu Thr Leu Phe

Signal peptide    +1    +1    Mature  $\alpha$ -trichosanthin  
 CTA ACA ACT CCT CCT GTG GAG GGC | GAT GTT AGC TTC CCT TTA TCA  
 Leu Thr Thr Pro Ala Val Glu Gly | Asp Val Ser Phe Arg Leu Ser

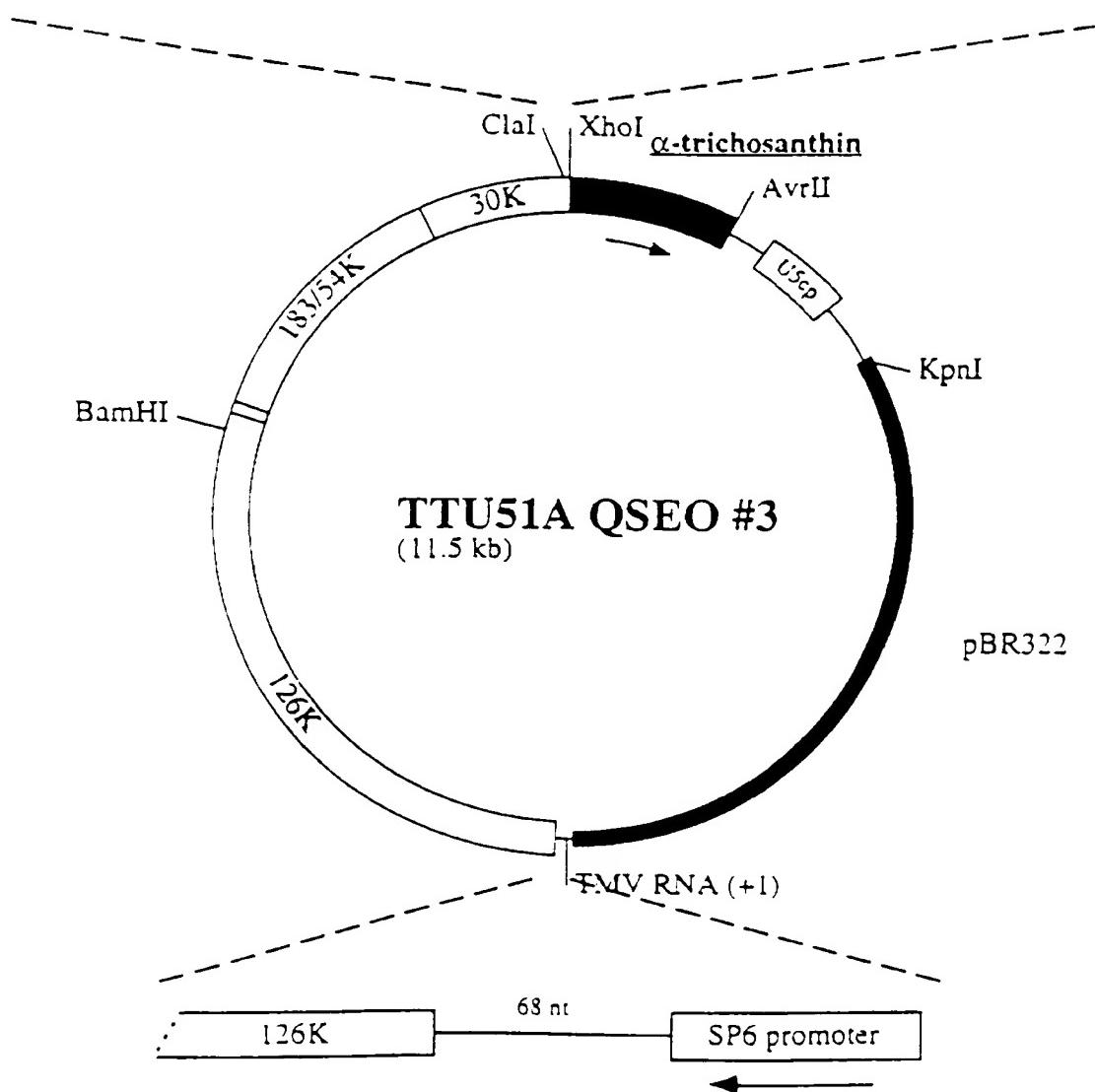
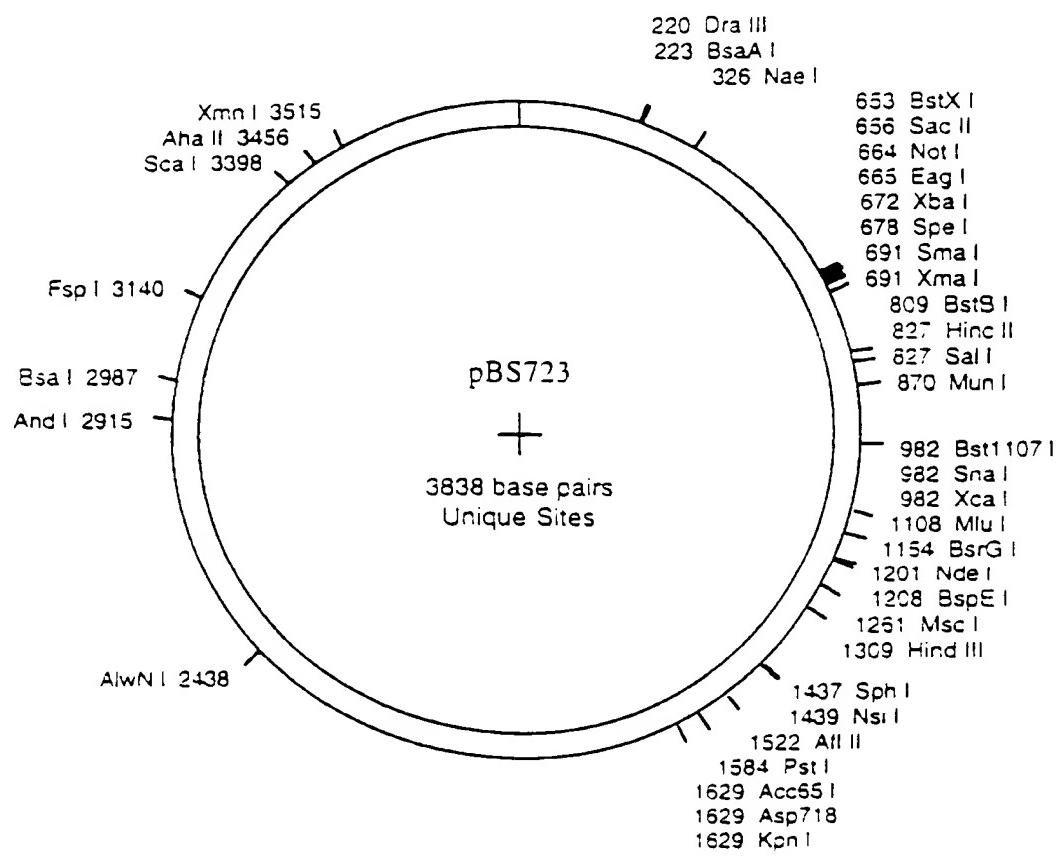
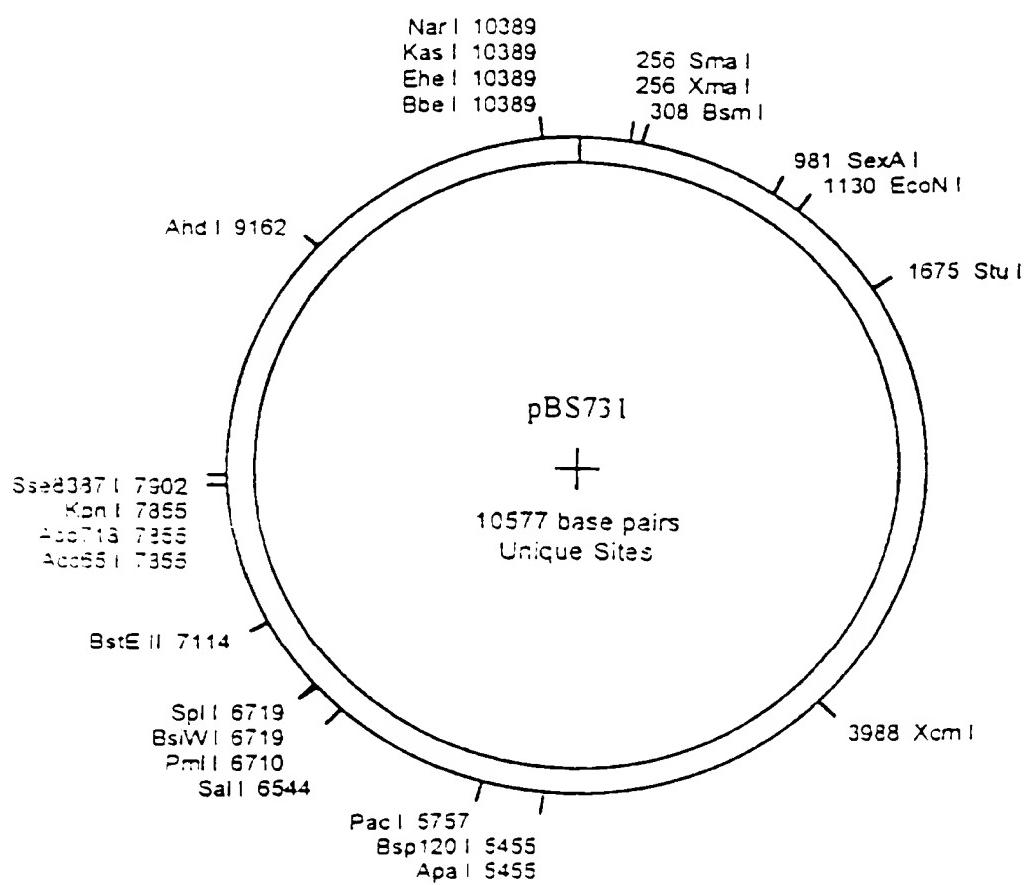


Figure 9



**Figure 10**



**Figure 11**

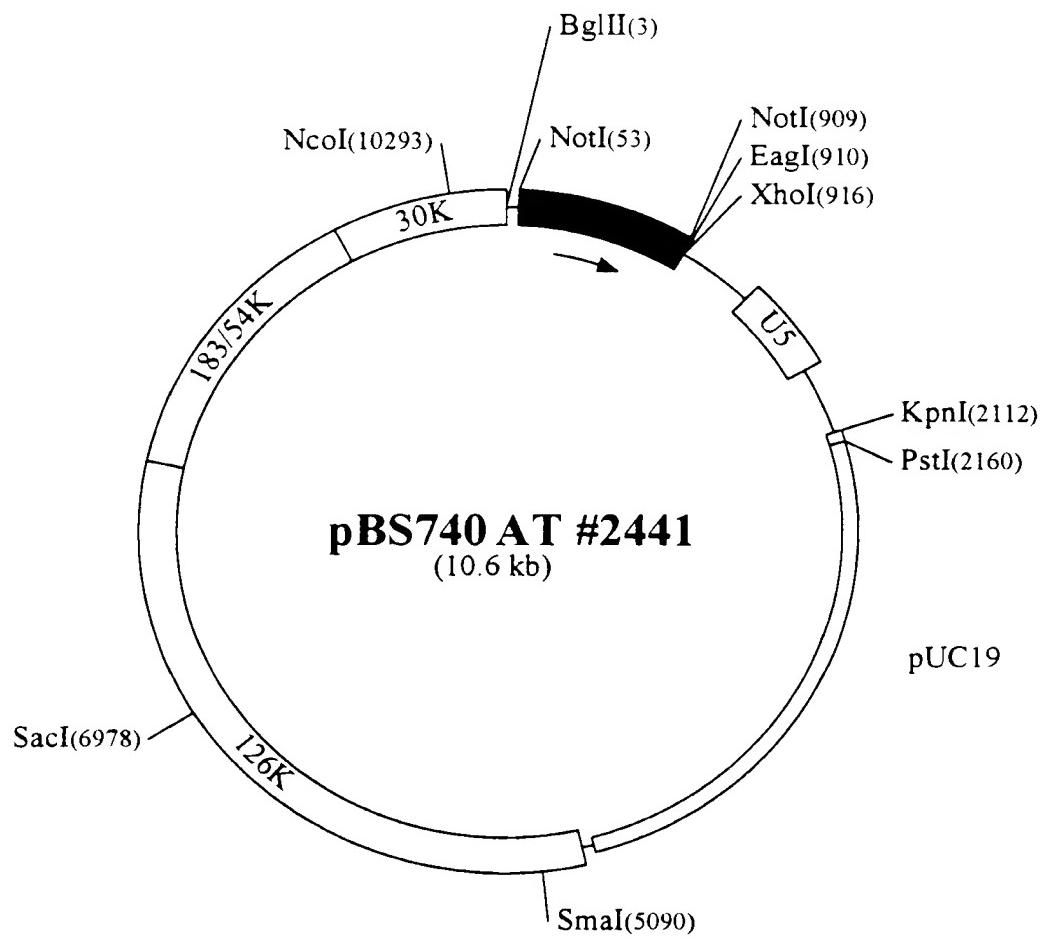


Figure 12

Nucleotide sequence of 740 AT #2441

ct tca ctt tcg ccg  
ATG GCT CTA CCT AAC CAG CAA ACC GTG GAT TAC CCT AGC TTC AAG CTC GTT ATC GTT GGC  
M A L P N Q Q T V D Y P S F K L V I V G  
GAT GGA GGC ACA GGG AAG ACC ACA TTT GTA AAG AGA CAT CTT ACT GGA GAG TTT GAG AAG  
D G G T G K T T F V K R H L T G E F E K  
AAG TAT GAA CCC ACT ATT GGT GTT GAG GTT CAT CCT CTT GAT TTC TTC ACT AAC TGT GGC  
K Y E P T I G V E V H P L D F F T N C G  
AAG ATC CGT TTC TAC TGT TGG GAT ACT GCT GGC CAA GAG AAA TTT GGT GGT CTT AGG GAT  
K I R F Y C W D T A G Q E K F G G L R D  
--  
GGT TAC TAC ATC CAT GGA CAA TGT GCT ATC ATC ATG TTT GAT GTC ACA GCA CGA CTG ACA  
G Y Y I H G Q C A I I M F D V T A R L T  
TAC AAG AAT GTT CCA ACA TGG CAC CGT GAT CTT TGC AGG GTT TGT GAA AAC ATC CCA ATT  
Y K N V P T W H R D L C R V C E N I P I  
GTT CTT TGT GGG AAT AAA GTT GAT GTG AAG AAC AGG CAA GTC AAG GCC AAG CAG GTA ACA  
V L C G N K V D V K N R Q V K A K Q V T  
TTC CAC AGG AAG AAC CTC CAG TAT TAC GAG ATA TCT GCC AAG AGC AAC TAC AAC TTC  
F H R K K N L Q Y Y E I S A K S N Y N F  
GAG AAG CCA TTC TTG TAC CTT GCT AGA AAG CTC GCC GGG GAC GCT AAT CTT CAC TTT GTG  
E K P F L Y L A R K L A G D A N L H F V  
GAA TCA CCT GCC CTT GCT CCC CCG GAA GTT CAA ATC GAC TTG GCT GCT CAG CAG CAG CAT  
E S P A L A P P E V Q I D L A A Q Q Q H  
GAG gCG GAG CTT GCA GCA GCA GCA AGT CAG CCA CTT CCT GAT GAC GAT GAT GAC ACC TTC  
E A E L A A A A S Q P L P D D D D T F  
GAG TAG AGA AAG AGA GAT GTG ATC TGT CAC TGA TTA CCC GTT AGG GCT TGT CTG AAC TTT  
E  
TTT TTG TTC ATG GTG CTA TTT TTA TGT GTC CGT ACT TTG AAA TGA ATC GAT GAC ATT AGT  
AAT TTT CAT TTT TAA GTT TTT AAC TGT CGC TAT GAA AGT GAA AAC

Figure 13

Nucleotide sequence alignment of 740 AT #2441 to AF017991  
*A. thaliana* salt stress inducible small GTP binding protein Ran1

740 AT #2441	1	CTTCACTTTGCCGATGGCTCACCTAACCAACAGCAAACCGTGGATTACCCTAGCTTCAGC 60      .....     .....     .....     .....     .....     .....     .....
AF017991	67	CTTCACTTTGCCGATGGCTCACCTAACCAACAGCAAACCGTGGATTACCCTAGCTTCAGC 125      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	61	TGTTATCGTTGGCGATGGAGGCACAGGGAAAGACCACATTGTAAAGAGACATCTTACTG 120      .....     .....     .....     .....     .....     .....     .....
AF017991	127	TGTTATCGTTGGCGATGGAGGCACAGGGAAAGACCACATTGTAAAGAGACATCTTACTG 186      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	121	GAGAGTTTGAGAAGAAGTATGAACCCACTATTGGTGTGAGGTTCATCCTCTGATTCT 180      .....     .....     .....     .....     .....     .....     .....
AF017991	187	GAGAGTTTGAGAAGAAGTATGAACCCACTATTGGTGTGAGGTTCATCCTCTGATTCT 246      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	181	TCACTTAAGTGGCAAGATCCGTTCTACTGTGGGATACTGCTGGCCAAGAGAAATTG 240      .....     .....     .....     .....     .....     .....     .....
AF017991	247	TCACTTAAGTGGCAAGATCCGTTCTACTGTGGGATACTGCTGGCCAAGAGAAATTG 306   --  .....     .....     .....     .....     .....     .....     .....
740 AT #2441	241	GCGGCTTACGGGATGGTTACTACATCCATGGACAATGTGCTATCATCAGTTGATGTCA 300      .....     .....     .....     .....     .....     .....     .....
AF017991	307	GCGGCTTACGGGATGGTTACTACATCCATGGACAATGTGCTATCATCAGTTGATGTCA 366      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	301	CAGCACCGACTGACATACAAGAATGTCCAACATGGCACCGTGATCTTGCAGGGTTGTG 360      .....     .....     .....     .....     .....     .....     .....
AF017991	367	CAGCACCGACTGACATACAGGAATGTCCAACATGGCACCGTGATCTTGCAGGGTTGTG 426      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	361	AAAACATCCCATTGTTCTTGTGGAAATAAAGTTGATGTGAAGAACAGGCAAGTCAAGG 420      .....     .....     .....     .....     .....     .....     .....
AF017991	427	AAAACATCCCATTGTTCTTGTGGAAATAAAGTTGATGTGAAGAACAGGCAAGTCAAGG 486      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	421	CCAACCGAGTAACATCCACAGGAAGAACCTCCAGTATTACGAGATATCTGCCAAGA 480      .....     .....     .....     .....     .....     .....     .....
AF017991	487	CCAACCGAGTAACATCCACAGGAAGGAGGAACCTCCAGTATTACGAGATATCTGCCAAGA 546      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	481	GCAACTACAACCTCGAGAAGCCATTCTGTACCTTGCTAGAAAGCTGCCGGGACGCTA 540      .....     .....     .....     .....     .....     .....     .....
AF017991	547	GCAACTACAACCTCGAGAAGCCATTCTGTACCTTGCTAGAAAGCTGCCGGGACGCTA 606      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	541	ATCTTCACTTGTGGAAATCACCTGCCCTTGCTCCCCCGGAAGTCAAATCGACTTGGCTG 600      .....     .....     .....     .....     .....     .....     .....
AF017991	607	ATCTTCACTTGTGGAAATCACCTGCCCTTGCTCCCCCGGAAGTCAAATCGACTTGGCTG 666      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	601	CTCACCGAGCATGAGGCCGGAGCTTGCACGCCAGCAAGTCAGCCACTTCTGATGACG 660      .....     .....     .....     .....     .....     .....     .....
AF017991	667	CTCACCGAGCATGAGGCCGGAGCTTGCACGCCAGCAAGTCAGCCACTTCTGATGACG 726      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	661	ATGATGACACCTTCGAGTAGAGAAAGAGAGATGTGATCTGCACTGATTACCGTTAGGG 720      .....     .....     .....     .....     .....     .....     .....
AF017991	727	ATGATGACACCTTCGAGTAGAGAAAGAGAGATGTGATCTGCACTGATTACCGTTAGGG 786      .....     .....     .....     .....     .....     .....     .....
740 AT #2441	721	CTTGTCTGAACCTTTTTT 738      .....     .....
AF017991	787	CTTGTCTGAACCTTTTTT 804      .....     .....

Figure 14

Nucleotide sequence alignment of 740 AT #2441 to L16787  
*N. tabacum* small ras-like GTP-binding protein

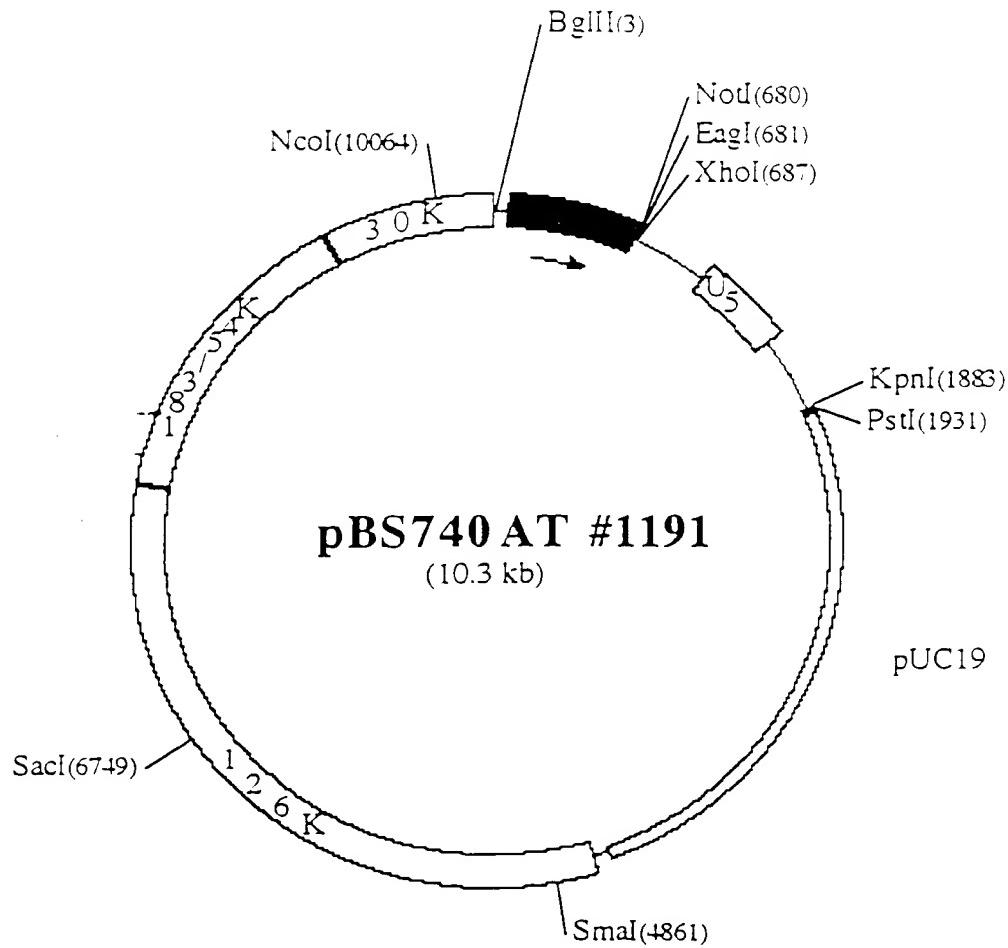
740 AT #2441	15	ATGGCTCTACCTAACCAAGCAAACCGTGGATTACCCTAGCTTCAGCTCGTTATCGTTGGC 74
L16787	37	ATGGCTCTACCAAAACCAACAAACTGTAGATTATCCAAGCTTCAGCTGTAATCGTGGGC 95
740 AT #2441	75	GATGGAGGCACAGGGAGACCAACATTTGTAAAGAGACATCTTACTGGAGAGTTTGAGAAG 134
L16787	97	GATGGAGGAACCTGGAAAACAACATTGTCAAGAGGCATCTTACTGGTGAATTGAGAAG 156
740 AT #2441	135	AAGTATGAACCCACTATTGGTGTGAGGTTCATCCTCTTGATTTCTTCACTAACTGTGGC 194
L16787	157	AAATATGAACCCACTATTGGTGTGGAGGTTCATCATTAGACTCTTACAAATTGTGGG 216
740 AT #2441	195	AAGATCCGTTCTACTGTGGGATACTGCTGGCAAGAGAAAATTGGTGGCTTAGGGAT 254
L16787	217	AAAATTGCTTTATTGCTGGGATACTGCTGGACAAGAGAAAGTTGGAGGTCTCGGGAT 276
740 AT #2441	255	GGTTACTACATGGACAAATGTGCTATCATCATGTTGATGTACAGCACCCAGACTGACA 314
L16787	277	GGTTACTACATTGCGCAATGCGCAATTATCATGTTGATGTACAGCCCCTGACC 336
740 AT #2441	315	TACAAGAATGTCACACATGCCACCGTGATTTGCAAGGGTTTGTGAAAACATCCCCATT 374
L16787	337	TACAAGAATGTCCTACGTGGCATCGAGATCTCTGCAGGGTTTGTGAAAACATCCCCATT 396
740 AT #2441	375	GTTCCTTGTGGAATAAAGTTGATGTGAAGAACAGGCAAGTCAGGCCAACCAAGTAACA 434
L16787	397	GTTCCTTGTGGAACAAAGTTGATGTCAAGAACAGGCAAGGTTAAGGCAAACCAAGTAC 456
740 AT #2441	435	TTCCACAGGAAGAACCTCCAGTATTACGAGATATCTGCCAAGAGCAACTACAACITC 494
L16787	457	TTCCACAGGAAGAAAAATTGCAATACTATGAGATCTCAGCAAAGAGTAACACTAACTTT 516
740 AT #2441	495	GAGAACCCATTCTTGTACCTTGCTAGAAAGCTGCCGGGGACGCTAACCTTCACTTGTG 554
L16787	517	GAGAACCTTTCTGTACCTTGCCAGAAAGCTTGCTGGGATGCTAACCTTCACTTGTG 576
740 AT #2441	555	GAATCACCTGCCCTTGCTCCCCCGGAAGTCAAATCGACTTGGCTGCTCAGCAGCAGCAT 614
L16787	577	GAATCACCTGCCACTTGCTCCCCCTGAAGTACAACATTGATTTAGCTGCACAGCAACTGCAT 636
740 AT #2441	615	GAGGGGGACCTTGCAAGCAGCAGCAAGTCAGGCCACTTCCGTGACGATGATGACACCTTC 674
L16787	637	GAACAAGACCTTTGCAAGCCGCTGCGCACGCACCTCCAGATGACGATGATGAGCTTTT 696
740 AT #2441	675	GAGTAGA 681
L16787	697	GAATAGA 703

Figure 15

Amino acid sequence comparison of 740 AT #2441 to tobacco RAN-B1 GTP binding protein

Nt RAN-B1	MALPNQQTVDYPSFKLVIVGDDGTGKTFVKRHLTGEFEKKYEPTIGVEVHPLDFFTNCG
740 AT #2441	
Nt RAN-B1	KIRFYCWDTAGQEKFGLRDGYIHGQCAIIMFDVTSTTDIQECSNMAP*SLQGL*KHSQ
740 AT #2441	+ +   + ++
Nt RAN-B1 <sup>-</sup>	LFFVGIKLM*KNRQVKAK
740 AT #2441	+   +       +
	IVLCGNKVDVKNRQVKAK

Figure 16

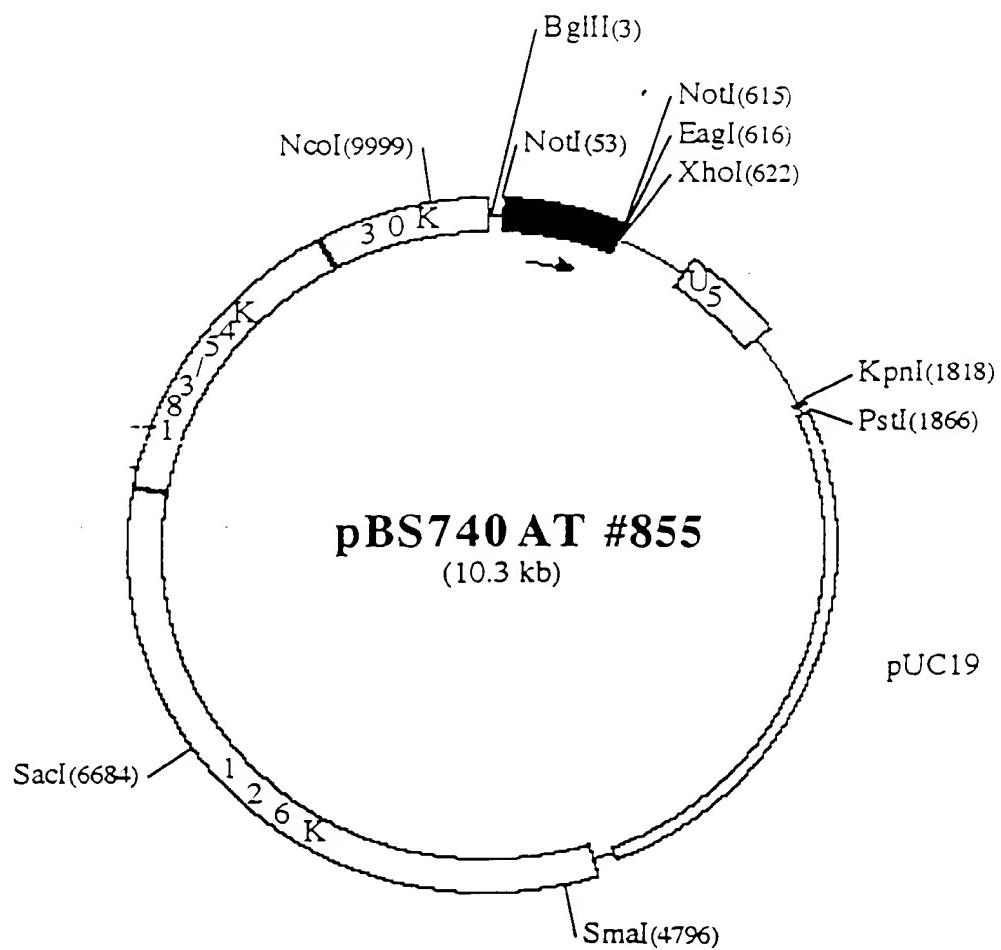


**Figure 17**

Nucleotide sequence of 740 AT #1191

GCT ACT ATG GTT GCC TCT CCG GCT CAG GCC ACT ATG GTC GCT CCT TTC AAC GGA CTT AAG  
A T M V A S P A Q A T M V A P F N G L K  
  
TCC TCC GCT GCC TTC CCA GCC ACC CGC AAG GCT AAC AAC GAC ATT ACT TCC ATC ACA AGC  
S S A A F P A T R K A N N D I T S I T S  
  
AAC GGC GGA AGA GTT AAC TGC|ATG CAG GTG TGG CCT CCG ATT GGA AAG AAG AAG TTT GAG  
N G G R V N C |M Q V W P P I G K K K F E  
  
ACT CTC TCT TAC CTT CCT GAC CTT ACC GAT TCC GAA TTG GCT AAG GAA GTT GAC TAC CTT  
T L S Y L P D L T D S E L A K E V D Y L  
  
ATC CGC AAC AAG TGG ATT CCT TGT GTT GAA TTC GAA GT  
I R N -K W I P C V E F E V

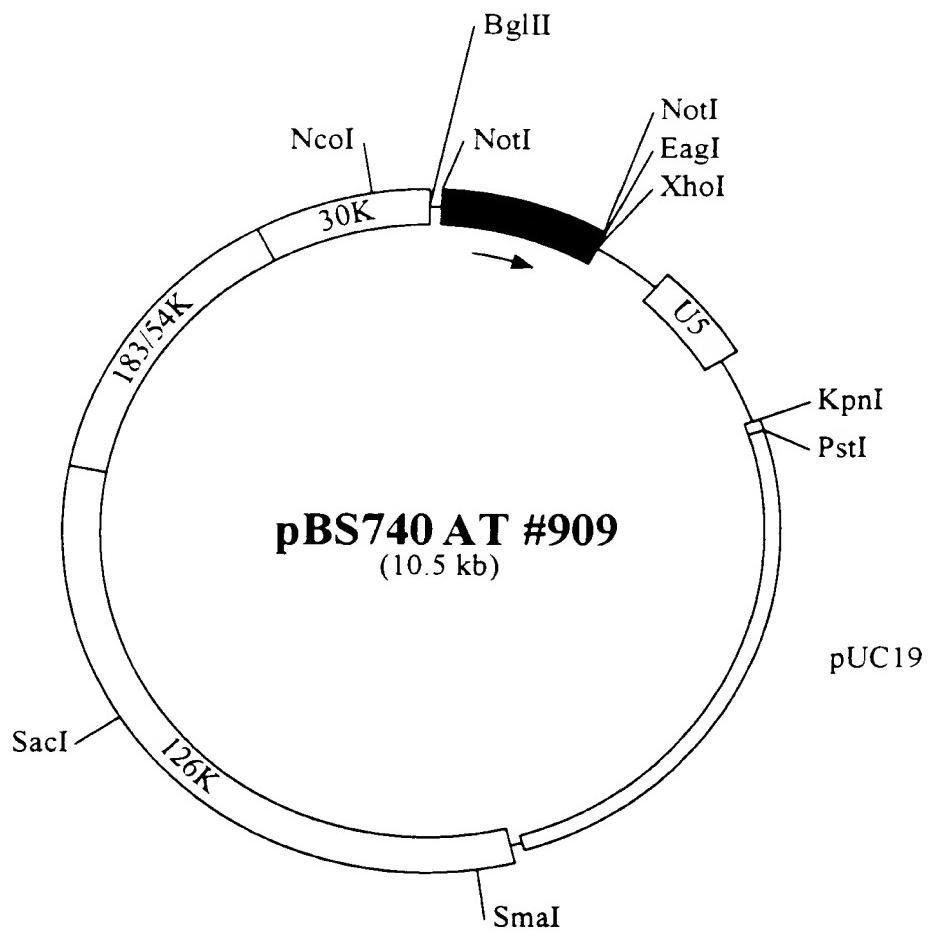
Figure 18



**Figure 19**

Nucleotide sequence alignment of 740 AT #855 to *Arabidopsis thaliana*  
HAT7 homeobox protein ORF (U09340)

**Figure 20**



**Figure 21**

**Nucleotide sequence alignment of 740 AT #909 to *H. sapiens* S56985 ribosomal protein L19 [human, breast cancer cell line, MCF-7]**

**Figure 22**

**Amino acid sequence alignment of 740 AT #909 to human P14118 60S ribosomal protein L19**

740 AT #909	1	KRLAASVMCGKGKVWLDPNESSDISMANSRQNIRKLVKDGFIIRKPTKIHRSRARKMK	60
		KRLA+SV++CGK KVWLDPNE+++I+ ANSRQ IRKL+KDG IIRKP +HSR+R RK	
Human P14118	8	KRLASSVLRCGKKVWLDPNETNEIANANSRQQIRKLIKDG LIIRKPVTVHSRARC KNT	67
740 AT #909	61	IAKMKGRHSGYGRKGTREARLPTKVLWMRRMRVLRLKKYRETKKIDKHYHDMYMRV	120
		+A+ KGRH G GKRKGTR AR+P KV WMRRMR+LRLLL++YRE+KKID+HMYH +Y++V	
Human P14118	68	LARRKGRHMGIGKRKGTA NARMPEKV TW MRRMRILRLLL RRYRESKKIDRHMYHSLYLV	127
740 AT #909	121	KGNVFKNKRVLMESIHKSA*KLGEK	146
		KGNVFKNKR+LME IHK KA K +K	
Human P14118	128	KGNVFKNKRILMEHIHKLADKARKK	153

**Figure 23**